Matter, M., Nenoff, L., Meier, G., Weber, D. C., Lomax, A. J., & Albertini, F. (2018). Alternatives to patient specific verification measurements in proton therapy: a comparative experimental study with intentional errors. *Physics in Medicine and Biology*, 63(20), 205014 (12 pp.).
<https://doi.org/10.1088/1361-6560/aae2f4>

Alternatives to patient specific verification measurements in proton therapy: A comparative experimental study with intentional errors

M Matter^{1,2}, L Nenoff^{1,2}, G Meier¹, D C Weber^{1,3,4}, A J Lomax^{1,2} and F Albertini¹

1 Paul Scherrer Institut, Center for Proton Therapy, Villigen PSI, Switzerland

2 Department of Physics, ETH Zürich, ETH Hoenggerberg, Zürich, Switzerland

3 Department of Radiation Oncology, University Hospital Zürich, 8091 Zürich, Switzerland

4 Department of Radiation Oncology, University Hospital Bern, 3000 Bern, Switzerland

E-mail: michael.matter@psi.ch

Abstract.

Patient specific verification (PSV) measurements for pencil beam scanning (PBS) proton therapy are resource-consuming and necessitate substantial beam time outside of clinical hours. As such, efforts to safely reduce the PSV-bottleneck in the clinical work-flow are of great interest. Here, capabilities of current PSV methods to ensure the treatment integrity were investigated and compared to an alternative approach of reconstructing the dose distribution directly from the machine control- or delivery log files with the help of an independent dose calculation (IDC). Scenarios representing a wide range of delivery or work-flow failures were identified (e.g. error in spot position, air gap or pre-absorber setting) and machine files were altered accordingly. This yielded 21 corrupted treatment files, which were delivered and measured with our clinical PSV protocol. IDC machine- and log file checks were also conducted and their sensitivity at detecting the errors compared to the measurements. Although some of the failure scenarios induced clinically relevant dose deviations in the patient geometry, the PSV measurement protocol only detected 1 out of 21 error scenarios. However, 11 and all 21 error scenarios were detected using dose reconstructions based on the log and machine files respectively. Our data suggests that, although commonly used in particle therapy centers, PSV measurements do a poor job detecting data transfer failures and imperfect delivery machine performance. Machine- and log-file IDCs have been shown to successfully detect erroneous work-flows and to represent a reliable addition to the QA procedure, with the potential to replace PSV.

Keywords: Proton therapy, Quality assurance (QA), Patient specific verification (PSV)
Submitted to: *Phys. Med. Biol.*

1. Introduction

In radiation therapy, uppermost importance is attached to guaranteeing the safety of the patient and the integrity of the treatment [Ford et al., 2012]. A combination of thorough commissioning, regular quality assurance (QA) procedures [Actis et al., 2017] and, for highly conformal therapy (e.g. proton therapy), patient specific verification (PSV) steps [Lomax et al., 2004a, Zhu et al., 2011, Furukawa et al., 2013, Li and Hsi, 2017, SSR, 2007] are implemented. This serves to assure flawless performance of the delivery machine, the treatment planning system (TPS) and the treatment work-flow. But it is a highly time consuming and resource demanding process. Especially the PSV measurements impose a substantial workload and, in the case of particle therapy, require much of the highly demanded beam time. Therefore efforts for safely removing this bottleneck in the clinical work-flow are of great interest [Zhu et al., 2015, Li et al., 2013, Mackin et al., 2013, Trnková et al., 2016, Meier et al., 2015].

Moreover, in recent years an increasing interest in the field of adaptive therapy is observed [Bernatowicz et al., 2018, Stock et al., 2017, Raaymakers et al., 2017]. In an online adaptive protocol, where the plan is adapted on the basis of an in-room CT or an MRI in treatment position, a PSV measurement is not possible [Noel et al., 2014]. Therefore the need for a fast and measurement free verification protocol is not only demanded by work-flow efficiency requirements but also driven by newly emerging technologies. Additionally, concerns questioning the effectiveness of PSV measurements for IMRT treatments have been raised [Ford et al., 2012] and the use of Gamma pass rate as a PSV measurement metric has been strongly criticized since it does not predict clinically relevant patient dose errors [Nelms et al., 2011, Carrasco et al., 2012].

The standard procedure currently adopted by most of the proton therapy facilities is to measure every single patient field before the delivery of the first fraction in a water phantom [Lomax et al., 2004a, Zhu et al., 2011, Furukawa et al., 2013, Li and Hsi, 2017]. These PSV measurements represent an end-to-end test of the work-flow steps between the accepted plan, and the treatment. The goal is to ensure agreement of the delivered dose to the plan, which is approved by a radiation oncologist. The measurement serves to check three key steps:

- Analytical dose calculation algorithm in water
- Data transfer from the TPS to the delivery machine
- Capability of the delivery machine to administer the adequate dose with the required accuracy

PSV measurements can detect failures originating from all three sources. As a 3D dose measurement of the entire dose distribution is typically not feasible for PSV, only a limited number of measurement points in the middle of the field are measured or a 2D measurement in a plane perpendicular to the beam is obtained. However, this measurement point sampling approach used in PSV is limited as it does not allow to detect all undesired dose deviations occurring away from those measurement points.

In the past years, second-check or independent dose calculation (IDC) engines have been introduced and are used in combination or as replacement of PSV measurements [Meier et al., 2015, Mackin et al., 2013]. IDCs allow for double checking of the dose calculation and, depending on the use-case, the data transfer between the TPS and the delivery machine [Lomax et al., 2004b]. An additional method emerging in the particle therapy QA tool-box is the analysis of the log-files, created during the delivery to report the machine performance [Li et al., 2013, Scandurra et al., 2016, Belosi et al., 2017]. Log-file analysis has been shown to enable dose reconstruction precise enough for PSV applications [Meier et al., 2015, Li et al., 2013] and may further simplify the PSV procedure if applied purposefully. With these new tools, particle therapy centers face the challenge of combining those methods to generate a second generation PSV, which optimizes work-flow efficiency without compromising patient safety. Some centers have started to take steps in this direction by deploying IDC dose reconstitution, allowing them to securely decrease the number of measurement points of every field [Zhu et al., 2015]. However, before PSV measurements can be more widely substituted, it is necessary to assess the performance of these alternatives in terms of reliability and accuracy in order to guarantee that they are at least as good as the measurements at ensuring the plan quality.

In this work, we have therefore conducted a failure detection sensitivity comparison between PSV measurements and IDC checks based on machine control- and log-files. The procedure and the paper structure is shortly outlined as follows. At first our delivery system, the clinical PSV measurement protocol and the IDC are described. IDC QA checks were defined and clinically applicable thresholds were obtained. Then possible work-flow or delivery failure scenarios were determined and their influence on the dose distribution was calculated and displayed in section 3.1. Those failures were artificially implemented in the treatment work-flow by altering machine files, simulating data corruption. Clinical PSV measurements were obtained and the newly defined QA checks were conducted. For the first time, failure detection sensitivity of different QA checks for realistic error scenarios in proton therapy are compared (section 3.3). Lastly advantages and shortcomings of the described QA-checks and steps towards further reducing PSV measurements are discussed.

2. Materials and methods

At PSI the proton source for clinical treatment is a cyclotron with a beam energy of 250 MeV. Gantry 2, which is used clinically since November 2013, is a PBS Gantry with a upstream energy selection design and range between 70-230 MeV [Pedroni et al., 2004]. The scanning magnets are designed to allow for fast parallel double scanning of the beams with a width between 2.5-4.5 mm σ in air [Pedroni et al., 2011]. From here on, those pencil beams will be referred to as spots. A range shifter of 4 cm water equivalent thickness can automatically be positioned into the beam path allowing for changing the pre-absorber setting of individual spots during the application of a field.

With the pre-absorber the beam width increases to 1 cm σ in air. Patient plans are generated with an in house developed TPS based on an analytical ray casting algorithm [Lomax et al., 1996]. In current clinical practice, and for every patient field, a PSV measurement is conducted. This procedure, as well as the IDC, which allows for other QA checks, are described in the following sections.

2.1. Clinical patient specific verification measurement protocol

PSV measurements are conducted using an in-house developed accordion-type water phantom with a two dimensional array of ionization chambers (27x27 chambers, seven29, PTW Freiburg, Germany) [Chowdhuri Z et al., 2016]. The water depth is variable and the phantom can be rotated to measure the patient field under oblique angles. The 729 vented plane-parallel ion chambers have a size of 5 mm x 5 mm x 5mm and a center-to-center spacing of 10 mm. The water phantom is positioned with the help of two orthogonal positioning lasers, which intersect at the center of the ionization chamber array. Before treatment, every patient field is measured at one depth. The measurement depth is chosen to be at the field center or, in the case of a multiple field plan, plains distributed around the field center are chosen. For example in a three field plan, one field is measured at the field center, one a couple of centimeters superficially and one distally. After the field application the measured data is read back to the TPS, where it is compared to the predicted dose distribution in water as calculated by the TPS. The evaluation method used at PSI, as described in previous publications [Lomax et al., 2004a, Trnková et al., 2016], is briefly outlined.

A setup error of the water phantom, an uncertainty in range and absolute dose need to be considered and quantified. Therefore, lateral and range shifts as well as a global dose scaling factor are applied, such that the standard deviation between the measurement and the dose prediction is minimized. This procedure yields four parameters (two lateral shifts, one range shift and one dose scaling factor), which together with the remaining standard deviation describe the agreement between the measurement and the calculation. The lateral shifts and the range shift only account for water phantom setup errors. The dose scaling factor describes the *absolute dose accuracy* and will henceforth be referred to as a dose discrepancy factor. If the calculated dose is larger than the measured dose, this results in a positive dose discrepancy factor. The remaining standard deviation describes the *relative dose accuracy*. A field has passed the verification if the four parameters as well as the standard deviation is within the tolerances as specified in table 1. Those tolerance criteria are derived from more than twenty years of clinical experience and are chosen such that a reasonable compromise between patient safety and work-flow efficiency is met, i.e. to minimize repeated measurements and/or replanning.

Table 1: Tolerance criteria for the clinical patient specific verification measurements.

	Tolerance
Dose discrepancy (% of prescription dose)	3%
Range precision	3 mm
Lateral precision	2 mm
Standard deviation (% of prescription dose)	4%

2.2. Independent dose calculation

The IDC, as described by Meier *et al* [Meier et al., 2015], is a QA toolkit, with which the dose distribution can be reconstructed from several input data streams, originating from different steps in the treatment work-flow. Different dose reconstruction algorithms are available in the frame work. Using the IDC, QA-checks can be designed aiming to examine specific steps in the treatment work-flow chain. For checking the analytical dose deposition algorithm a Monte Carlo engine is the appropriate tool. On the other hand, for checking data transfer integrity and delivery machine performance, using the same ray casting algorithm, the same water equivalent depth algorithm and the same beam data as the TPS allows for exactly pinpointing the origin of observed differences in dose. Therefore the later approach was chosen for the two IDC dose reconstruction configurations, which are used as QA-tools in this paper and are outlined in the following paragraphs. The problem of individually checking the dose algorithm will be picked up again in the discussion.

2.2.1. Machine file based dose reconstruction To apply a patient treatment, the delivery machine needs all the data describing the patient fields. To that end, a machine file is generated. The machine file contains all field and spot information as well as extensive additional machine parameters describing the treatment. The data transfer from the TPS data into the machine file involves non trivial data conversions e.g. spot sorting, energy layer initialization and number of proton to monitor unit count conversions. The machine file based dose reconstruction uses this data to calculate the dose in the patient geometry with the same algorithm as the TPS. However the implementation of the dose calculation is independent of the TPS (written in a different programming language) and the beam data is accessed from a different source. This allows for a redundant check of the TPS dose calculation, as well as the validity of the machine file generation. Dose distributions calculated with the machine file based dose reconstruction display differences to the nominal dose distribution of up to 0.3% of the prescription dose [Meier et al., 2015]. This disagreement arises from differences in treatment of floating point numbers and rounding errors when transforming between number of protons and monitor units. The high agreement of the IDC and the TPS allows for detecting very small differences in the input parameters and is therefore the appropriate tool to check the flawless generation of the machine file. As such, machine

file based dose reconstruction was used as a patient specific QA tool in this work, with a tolerance threshold of 0.4% in absolute voxel dose difference.

2.2.2. Log file based dose reconstruction During the application of a field the proton fluence is measured by two transmission chambers. Simultaneously the position of every pencil beam in the transversal plane is measured by strip monitors located in the gantry nozzle [Actis et al., 2014] with an accuracy of 0.14 mm [Scandurra et al., 2016]. This information with all other parameters and settings, relevant for the treatment, are recorded in real-time into a log file [Grossmann, 2007]. The log file based dose reconstruction uses this information as input data and calculates the dose in the patient geometry. With this method, the correct interpretation of the machine file by the delivery machine and the delivery machine accuracy are examined. In the absence of a treatment work flow error, discrepancies larger than 0.3% of prescription dose can be attributed to deviations in machine delivery performance. It is reported that the main source for deviations between the log file based dose reconstruction and the nominal TPS plan are deviations in spot positions [Scandurra et al., 2016]. To determine meaningful tolerance criteria, 507 patient fields from 106 patients treated between July 2015 and March 2018 were analyzed. This analysis included the data previously analyzed by [Scandurra et al., 2016] and [Belosi et al., 2017] and was extended by including more recent data and additional parameters. The largest voxel dose differences and standard deviations (ignoring voxels in air and voxels with dose below 10%) of all fields were calculated. Patient fields, which were generated correctly and deliver a clinically acceptable dose, should rarely fail the log file based dose reconstruction QA test, since this would otherwise unnecessarily complicate clinical QA procedure. For this paper, the tolerance limit for absolute voxel dose difference was chosen to be 10% and for the standard deviation 1.5%. Only 2 out of the 507 investigated patients exceeded the chosen largest voxel dose difference threshold and 6 the standard deviation threshold. With these tolerance limits only 1.4% of the investigated patient fields would have failed the proposed QA-check.

2.3. Investigated patient field

All further investigations were conducted on a single IMPT field (figure 1 (right)) of a clinically realistic four field plan of a head and neck patient with a tumor in the nasopharyngeal region (figure 1 (left)). For this plan all spots were delivered with a pre-absorber. The abundance of density heterogeneities in the beam path (nasal cavities) as well as the high dose modulation of the IMPT plan, make this a good example of a worst case scenario. It is not the goal of this study to determine the magnitude of expected dose deviations given a certain error scenario, as this depends strongly on the investigated field. The objective is rather to show advantages and limitations of QA measurements and QA tools, as well as comparing their detection sensitivity for a given dose deviation by investigating a single worst case field.

Figure 1: Investigated patient treatment with indication of PTV in red. Four field IMPT plan (left). The investigated dose field from the IMPT plan (right).

2.4. Investigation of spot and field information changes on dose distribution

Seven different alterations representing a wide range of possible work flow or delivery failures have been identified:

- *Shift one spot*
- *Remove one spot*
- *Remove pre-absorber for one spot*
- *Remove low weighted spots*
- *Increase air gap*
- *Shift all spots randomly*
- *Increase weights for all spots*

With the IDC, the dose distributions with these alterations were calculated in the patient geometry, to get a better understanding of their effect on the dose. All alterations were examined separately, with combinations not being considered. For every alteration scenario twenty different severities (except for the air gap alteration only ten severities) were calculated. The dose distribution differences were quantified with three metrics:

- Largest absolute voxel dose difference (MaxDiff) given in % of prescription dose
- Standard deviation of all voxels with dose above 10% and Hounsfield unit (HU) above threshold to ignore voxels in air given in % of prescription dose
- Gamma test fail rate for pass criteria 3% / 3 mm considering all voxels with dose above 10% and HU above threshold (-994) to ignore voxels in air given in % of prescription dose

The first three alterations (*Shift one spot*, *Remove one spot* and *Remove pre-absorber for one spot*) are changes on a single spot out of approximately nine thousand spots. The spot information like position, spot weight, energy and pre-absorber setting

of all spots is written in the machine file. An error in machine file generation might alter the spot information. Another possibility for such an error to originate would be if one spot is omitted because of a treatment interruption or if the position of a single spot is distorted due to delivery machine malfunction.

In the *Remove low weighted spots* alteration scenario between 50 and 1000 low weighted spots out of nine thousand spots were removed. Both, the TPS and the delivery machine, have a limit for minimum spot weight, which for the delivery machine depends on the proton current. If those two limits do not correspond, such an error might appear. The air gap is determined by the nozzle extension, a machine setting saved in the machine file. An increase in the air gap causes the spot width to increase and therefore the dose distribution to blur. This affects all the spots in the field as investigated in the *Increase airgap* alteration scenario. For the alteration *Shift all spots randomly* all spots of the field were laterally shifted in both orthogonal directions to the beam direction by adding a different normally distributed error with a sigma between 0 and 6 mm. This simulated an imprecise delivery machine. A scaling of all spots as for *Increase weights for all spots* might come from a wrong dose to monitor unit calibration or a false manual absolute dose adjustment, which some centers do to correct their dose algorithm or delivery machine characteristics.

2.5. Alter machine files for comparing PSV measurements and IDC QA-tools

The machine file of the nominal field was altered, simulating corruption of data transfer or parameter transformation between the TPS and the delivery machine. For each of the seven alteration scenarios, for which the effects were previously calculated (section 3.1), two to four severities were chosen and the corresponding changes were applied, yielding 21 altered machine files. A listing of the different machine file alterations is given in table 2. Our clinical PSV measurement protocol was applied on those files and the two described IDC QA-checks were conducted.

3. Results

3.1. Investigation of spot and field information changes on dose distribution

The results for the three alterations affecting single spots are given in figure 2. Alterations on single spots affect only a small volume of the field. Therefore the standard deviation and the Gamma fail rate are respectively small. Categorically the difference metrics rise for an increase in alteration severity. In the *Remove one spot* and *Remove pre-absorber for one spot* alteration scenarios the evoked dose deviations do not only depend on the spot weight, but also on the position of the spot in the field. This is the cause for the steps in the curve. The results for the four alteration scenarios affecting the whole field are given in figure 3. In all cases the difference parameters rise for increasing alteration severities. Also it can be clearly observed that the Gamma fail rate only starts to increase for spot random position errors with a sigma above 1.5 mm

Table 2: Machine file alterations generated for patient specific verification QA test

Alteration	Severity	Description
Shift one spot	1 mm	The highest weighted spot is shifted by 1 mm in lateral direction
	3 mm	
	5 mm	
	10 mm	
Remove one spot	5%	One spot is removed with a weight 5% of max weight
	10%	
	60%	
Remove pre-absorber for one spot	5%	The pre-absorber is removed for one spot with a weight 5% of max weight
	10%	
	60%	
Remove low weighted spots	0.7%	The 0.7% lowest weighted spots are removed
	2.5%	
	5.5%	
	12%	
Increase air gap	1 cm	The air gap is increased by 1 cm
	2 cm	
	5 cm	
Shift all spots randomly	1 mm	All spots are shifted randomly in both lateral directions with sigma of 1 mm
	2 mm	
Increase weights for all spots	3%	Increase the weights of all spots by 3%
	5%	

or global weight increase of 3%. Two examples of dose error distributions for alteration scenarios with a high impact on the dose are shown in figure 4. The dose difference caused by removing the pre-absorber of one spot with a relative weight of 60% of the maximum spot weight of this field displays a hot spot with 5% over dosage and a cold spot with 8% under dosage inside the PTV. The dose difference distribution caused by removing the 12% lowest weighted spots shows an under dosage of more than half the target by 5% or more. The alteration severities chosen to generate the corresponding altered machine files (see section 2.5) are indicated with vertical lines in figures 2 and 3.

3.2. Patient specific verification measurements of altered machine files

The results of the clinical PSV water phantom measurement are given in table 3. For all measurements a setup correction of 1 mm in one lateral direction was applied. In the

Figure 2: Deviation in the dose distribution calculated in the patient geometry caused by single spot information changes of different severities. The alteration severities chosen to generate the corresponding altered machine files are indicated with vertical lines. The Gamma fail rate is indicated by squares, the standard deviation by circles and the maximal voxel difference by triangles.

other lateral direction no correction was necessary. The measurement depth was chosen to be at the field center, which corresponded to 5.5 cm depth in water. The nominal machine file, as well as all altered files except one, passed the PSV measurement check. The analysis method of the water phantom measurements is designed to detect errors in *absolute dose accuracy* with a high sensitivity. This is confirmed by the results. For the machine files with the *Increase weights for all spots* alteration the increase was correctly determined as can be seen by the matching dose discrepancy factor. For the case where all the weights were increased by 5% the measured dose discrepancy exhibited 5% dose surplus in the measurements, which is out of the tolerance clinically used at PSI (table 1) and therefore the PSV measurement test failed. For all other machine file alteration scenarios the measurement parameters dose discrepancy factor, shift in range, lateral shifts and standard deviation were inside the tolerance limits and therefore the PSV measurement check is passed.

Even though clinically the verification measurement is only conducted in one depth, for the machine file alterations, where only one spot is affected, the measurements were repeated at a shallow depth of 1.5 cm in water. The reason why this shallow depth was chosen is that for the investigated field the highest weighted spots are at this shallow position (see figure 1). For single field uniform dose plans the highest weighted spots

Figure 3: Deviation in the dose distribution calculated in the patient geometry caused by field information changes of different severities. The alteration severities chosen to generate the corresponding altered machine files are indicated with vertical lines. The Gamma fail rate is indicated by squares, the standard deviation by circles and the maximal voxel difference by triangles.

Figure 4: Calculated dose difference caused by removing the pre-absorber of a spot with 60% of maximal spot weight (left) and due to removing the 12.5% lowest weighted spots (right). The PTV is outlined in red.

Table 3: Patient specific verification measurement of altered machine files measured in the field center. Only the highest severity of the alteration scenario *Increase weights* for all spots displays a value out of tolerance as indicated with a red font.

Alteration	Severity	Dose discrepancy	Shift range	MaxDiff	STD	QA test
No alteration		0%	0 mm	5.7%	1.0%	pass
Shift one spot	1 mm	-1%	0 mm	4.1%	1.3%	pass
	3 mm	0%	0 mm	4.9%	0.9%	pass
	5 mm	0%	0 mm	3.9%	0.9%	pass
	10 mm	0%	0 mm	6.2%	0.8%	pass
Remove one spot	5%	0%	0 mm	4.0%	0.9%	pass
	10%	0%	0 mm	4.3%	1.0%	pass
	60%	0%	0 mm	3.9%	0.9%	pass
Remove pre-absorber for one spot	5%	0%	0 mm	4.6%	1.1%	pass
	10%	0%	0 mm	4.9%	1.0%	pass
	60%	0%	0 mm	17.8%	1.4%	pass
Remove low weighted spots	0.7%	0%	0 mm	4.1%	1.0%	pass
	2.5%	0%	0 mm	4.9%	1.0%	pass
	5.5%	1%	0 mm	3.9%	1.0%	pass
	12%	2%	0 mm	6.2%	1.1%	pass
Increase air gap	1 cm	0%	0 mm	6.0%	1.1%	pass
	2 cm	1%	0 mm	8.0%	1.4%	pass
	5 cm	2%	0 mm	9.0%	1.8%	pass
Shift all spots randomly	1 mm	0%	0 mm	6.5%	1.3%	pass
	2 mm	1%	0 mm	11.4%	2.1%	pass
Increase weights for all spots	3%	-3%	0 mm	3.9%	1.0%	pass
	5%	-5%	0 mm	4.8%	1.0%	fail

are always close to the distal edge of the target due to the nature of the spread out Bragg peak. In IMPT plans however, due to the increased modulation this is not true anymore and superficial high weighted spots are common. An additional advantage of such a shallow measurement is that the dose distribution is made up of almost all spots of the field, so potentially errors from anywhere in the field could be detected. The measurement results for the PSV measurements at 1.5 cm depth for the machine file alterations affecting single spot information is given in table 4. The values for dose discrepancy, the range shift and the standard deviation are increased compared to the measurement at the field center. The reason for the increase of those values, also in the scenario where no alteration was introduced, is that the dose algorithm does not consider the build up effect. Nevertheless all measurement parameters are inside tolerance and therefore all fields passed the PSV measurement check also at this depth.

Table 4: Patient specific verification measurement of altered machine files measured at 1.5 cm water depth.

Alteration	Severity	Dose discrepancy	Shift range	MaxDiff	STD	QA test
No alteration		1%	-1 mm	7.9%	1.4%	pass
Shift one spot	1 mm	0%	0 mm	7.6%	1.8%	pass
	3 mm	1%	-1 mm	6.6%	1.6%	pass
	5 mm	1%	-1 mm	7.2%	1.6%	pass
	10 mm	1%	-1 mm	13.1%	2.0%	pass
Remove one spot	5%	0%	-1 mm	7.5%	1.7%	pass
	10%	-1%	-1 mm	7.1%	1.8%	pass
	60%	3%	0 mm	7.9%	1.8%	pass
Remove pre-absorber for one spot	5%	1%	-1 mm	6.5%	1.7%	pass
	10%	2%	-1 mm	6.2%	1.7%	pass
	60%	2%	-1 mm	6.8%	1.6%	pass

3.3. IDC QA-checks of altered machine files

The results of the machine- and log file IDC QA-check are given in table 5. As the dose calculation of the IDC matched those from the TPS, this allowed for the use of small tolerance limits for the machine file dose reconstruction test, without risk of false positive test outcomes. All induced changes caused a deviation in dose reconstruction, which exceeded the tolerance limit and the machine file QA test therefore correctly detected all machine file alteration. For the log file QA test the tolerance limits are higher, since machine delivery inaccuracies are included in the measurement. 11 out of 21 machine file alterations were detected including the highest severity of every alteration scenario. The detection sensitivity for alterations in the machine file of PSV measurements and the IDC QA-checks are compared in table 6.

4. Discussion

The authors have compared the failure detection sensitivity of PSV measurements and alternative patient specific QA methods, the latter not requiring extra beam time. This comparison was conducted on a single representative patient field, while the applied tolerance criteria for the suggested IDC QA checks were obtained reviewing treatment data of more than 500 applied patient fields. It has been shown that IDC dose reconstruction from machine control- and log-files have a higher sensitivity to detect faulty machine control files than PSV measurements. There are two reasons for the low error detection sensitivity of the investigated PSV measurement protocol. One is that the measurement is only obtained in one plane and the highest deviation caused by the machine file alterations might appear at another position in the dose field. Another reason is the use of the high tolerance limits, since for a lot of the alteration scenarios the

Table 5: Machine file and log file based dose reconstruction QA test results of all altered machine files. The values out of tolerance are depicted in red.

Alteration	Severity	Machine file reconstruction			Log file reconstruction		
		MaxDiff	STD	QA test	MaxDiff	STD	QA test
No alteration		0.3%	0.03%	pass	1.3%	0.37%	pass
Shift one spot	1 mm	3.0%	0.11%	fail	7.5%	0.78%	pass
	3 mm	8.7%	0.34%	fail	9.0%	0.47%	pass
	5 mm	14.0%	0.56%	fail	14.0%	0.68%	fail
	10 mm	26.5%	1.06%	fail	26.6%	1.28%	fail
Remove one spot	5%	1.9%	0.10%	fail	2.0%	0.23%	pass
	10%	3.4%	0.21%	fail	3.2%	0.32%	pass
	60%	21.9%	1.05%	fail	21.1%	1.09%	fail
Remove pre-absorber for one spot	5%	2.3%	0.07%	fail	2.1%	0.24%	pass
	10%	8.1%	0.17%	fail	8.1%	0.26%	pass
	60%	62.8%	0.96%	fail	62.8%	0.99%	fail
Remove low weighted spots	0.7%	1.17%	0.14%	fail	1.7%	0.39%	pass
	2.5%	3.8%	0.55%	fail	4.0%	0.55%	pass
	5.5%	9.8%	1.32%	fail	10.1%	1.33%	fail
	12%	17.2%	2.98%	fail	17.0%	2.93%	fail
Increase air gap	1 cm	6.1%	0.71%	fail	6.3%	0.84%	pass
	2 cm	15.9%	1.98%	fail	16.1%	2.20%	fail
	5 cm	24.5%	3.24%	fail	24.4%	3.49%	fail
Shift all spots randomly	1 mm	7.9%	1.55%	fail	8.3%	1.59%	fail
	2 mm	13.3%	3.18%	fail	13.5%	3.18%	fail
Increase weights for all spots	3%	4.8%	1.02%	fail	5.0%	1.08%	pass
	5%	8.0%	1.70%	fail	8.7%	1.86%	fail

dose discrepancy or the standard deviation is clearly elevated compared to the nominal case. For example, as reported in table 3, the highest level of severity in the introduced error scenario *Remove low weighted spots* and *Increase air gap* show a dose discrepancy of 2% and the scenario *Shift all spots randomly* displays a standard deviation above 2%. Thus the QA criteria are higher than for the PSV measurements in the nominal case (0% dose discrepancy and 1% standard deviation).

So why are the tolerance limits so high, that these alterations are not detected? The reason is related to the "end-to-end" test character of PSV measurements. The biggest advantage of being able to detect failures or dose deviations occurring anywhere in the work-flow chain between the plan and the delivery, also entails serious drawbacks. The tolerance limits need to include deviations from dose calculation inaccuracies rendering the QA check blind to errors in data transfer or delivery inaccuracies. If a deviation is detected, it is impossible to pinpoint the origin of the error in the work-

Table 6: Comparison of detection sensitivity to changes in the machine file of PSV measurements, IDC machine file and log file dose reconstruction QA test.

Alteration	Severity	PSV measurement	Machine file reconstruction	Log file reconstruction
No alteration		Passed	Passed	Passed
Shift one spot	1 mm	Missed	Detected	Missed
	3 mm	Missed	Detected	Missed
	5 mm	Missed	Detected	Detected
	10 mm	Missed	Detected	Detected
Remove one spot	5%	Missed	Detected	Missed
	10%	Missed	Detected	Missed
	60%	Missed	Detected	Detected
Remove pre- absorber for one spot	5%	Missed	Detected	Missed
	10%	Missed	Detected	Missed
	60%	Missed	Detected	Detected
Remove low weighted spots	0.7%	Missed	Detected	Missed
	2.5%	Missed	Detected	Missed
	5.5%	Missed	Detected	Detected
	12%	Missed	Detected	Detected
Increase air gap	1 cm	Missed	Detected	Missed
	2 cm	Missed	Detected	Detected
	5 cm	Missed	Detected	Detected
Shift all spots randomly	1 mm	Missed	Detected	Detected
	2 mm	Missed	Detected	Detected
Increase weights for all spots	3%	Missed	Detected	Missed
	5%	Detected	Detected	Detected

flow. Increasing the number of measurement planes or lowering the tolerance limits for PSV measurements would to some extent increase error detection sensitivity. This however would strongly increase the work-load for the already expensive PSV procedure and the problem of pinpointing the origin of the error would remain. Preferably a 3D measurement could be obtained, but to our knowledge no modality, which fulfills the accuracy demands and is easy and fast enough to be employed for PSV measurements exists.

The IDC enables more specified tests to check different steps separately, with a higher failure detection sensitivity. In this way it is possible to detect unexpected differences between the planned and the actually applied dose, which are much smaller than one would consider clinically relevant. Detecting those allows for spotting possible problems in the work-flow. Due to the restricted scope of the single tests, it is easy to locate the origin of an error and steps to prevent a reoccurrence can be implemented.

Even such small deviations, which pass Gamma pass rate checks or PSV measurement protocols as described here, might cause clinically relevant problems if accidentally the highest deviation falls on a critical organ at risk. In this work we have used similar QA criteria for the IDC QA checks as for the PSV measurements, which do not consider the underlying patient anatomy. As PSV measurements are performed in a water phantom, it is not possible to transfer the results into more clinically relevant parameters. However, since the IDC dose is reconstructed in the patient geometry, dose error effects on dose volume histograms can easily be calculated and for example acceptable maximum voxel differences depending on the underlying anatomical structure could be defined. The current PSV measurements cannot verify the steep dose gradients at the distal end of the dose field. Machine and log file reconstruction however, verify the steep gradients indirectly by relying on previously recorded depth dose curves, which are verified during the daily QA.

In summary, the error detection sensitivity of the investigated QA tests depends strongly on the applied tolerance limits. For all three methods it is a trade off between sensitivity and work-flow efficiency. All patient field QA tests, which result in a false positive, cause considerable extra clinical effort. For the clinically used and here applied PSV measurement protocol, the tolerance limits come from over 20 years of clinical experience. The newly defined IDC QA-check tolerance limits were defined, such that they have a similar failure probability for clinically applied fields in the past. We assume that for the 507 considered fields no work-flow, data transfer or delivery error occurred. Therefore, using the defined tolerance limits allow to make a fair comparison of error detection sensitivity between the established PSV measurements and the newly defined IDC QA-checks.

The machine file dose reconstruction obviously has the highest detection sensitivity of the compared QA-checks, since no delivery or dose calculation uncertainty needs to be considered. The drawback is that its detection span is quite limited, e.g. it can only detect specific errors in the dose calculation or during the transfer of plan information into the machine file. Nevertheless the authors believe that this test is an important QA step. While the machine file analysis is not so widely spread, the data conversion between the TPS data and the delivery machine data is a necessity. For commercial systems, typically the machine data is not extractable in the format of a machine file. However, if a separate QA-check wants to be conducted on this level, this data needs to be made available. For QA purposes this data should consist of pure delivery machine instructions, such that all upstream data conversions can be checked.

To answer the question, what is needed to replace PSV measurements, the three areas checked by PSV measurements are revised: analytical dose calculation algorithm in water, data transfer from the TPS to the delivery machine and capability of the delivery machine to administer the specific plan with the required accuracy. Data transfer failures can be detected by the machine file and the log file dose reconstruction check and delivery machine accuracy by the log file check. As already foreshadowed the analytical dose calculation algorithm can be checked by comparing

the calculation to a different algorithm [Raaymakers et al., 2017, Mackin et al., 2013]. For proton therapy the method of choice are Monte Carlo dose calculation engines [Grassberger et al., 2015, Grevillot et al., 2011]. Such an algorithm has recently been integrated in the IDC system in our institute. The calculated *absolute dose accuracy* of this implementation has been reported to agree within $\pm 2\%$ to measurements in the water phantom [Winterhalter et al., 2018]. This method allows for checking the analytical algorithm not only in water, but also in patient geometry. Addressing the advantages and problems of replacing measurements with Monte Carlo simulations for the analytical dose calculation algorithm check, although interesting and of high importance, exceeds the scope of this work. The two other areas (data transfer and delivery) are covered by the two discussed QA-checks, which are able to detect deviations with a higher sensitivity than PSV measurements.

There are remaining risks of replacing PSV measurements with IDC checks identified by the authors. Checking the delivery accuracy with the log file dose reconstruction, instead of the measurement, has some inherent problems. These are: (1) The log file records contain measurements of the beam only in the nozzle and not at the actual iso-center, (2) the measurement is not independent of the delivery machine, (3) the range is not measured and (4) dose is measured in monitor units and not in Gy. However, combining this analysis with an appropriate daily check, where the monitor units to Gy conversion and the beam range is measured, removes concerns (3) and (4). This procedure has recently been introduced at our institute [Actis et al., 2014]. The limitations (1) and (2) should be carefully considered during the commissioning and regular QA comparing log-file reconstructed doses to measured doses for patient fields should be implemented. One could think of a scenario where some patient fields are randomly chosen for standard PSV measurements, where comparing the measurements to the log-file dose is very little additional effort and serves as a log file dose reconstruction QA. The other patients' PSV would be conducted based on log files only. Another prominent risk of replacing PSV measurements with IDC checks arises from replacing the PSV measurement "end-to-end" test with the three described IDC QA-checks, which have a limited detection scope, there is always a risk of a failure scenario falling in-between the checks and not being detected.

For IMRT different centers have started to partially replace PSV measurements for standard patient fields with IDC software approaches and for the first patients treated with a MR-linac, a fast online measurement free QA assurance protocol has been successfully deployed [Raaymakers et al., 2017]. Work-flow efficiency demands are likely to push proton therapy in the same direction and online adaptive therapy protocols will demand a QA procedure without actually applying the patient field before the treatment. While dose calculation algorithm and data transfer checks can be conducted beforehand, the delivery accuracy check will be done retrospectively using the log files of the first fraction. In the case of unacceptable delivery accuracy during the first fraction one could think of correcting for this failure in the subsequent fractions, using the log-file based dose reconstruction as a feedback for the next fraction.

5. Conclusion

It is demonstrated that, although commonly used in many centers, PSV measurements are not able to guarantee the integrity of the data transfer and optimal delivery machine performance. It is shown that QA checks based on machine- and log file based dose reconstruction may do a better job in detecting failures in data transfer, transformations or delivery than PSV measurements.

6. Acknowledgement

This publication was formed in the scope of the Swiss national science foundation project "Towards the Daily Adapted Proton Therapy at PSI." The authors would like to acknowledge SNF for their support with the grant SNF: 320030_165961.

References

- [SSR, 2007] (2007). Quality control for Intensity-modulated radiation therapy. *Swiss Society for Radiobiology and Medical Physics*.
- [Actis et al., 2017] Actis, O., Meer, D., König, S., Weber, D. C., and Mayor, A. (2017). A comprehensive and efficient daily quality assurance for PBS proton therapy. *Phys. Med. Biol.*, 62.
- [Actis et al., 2014] Actis, O., Meer, D., and Onig, S. (2014). Precise on-line position measurement for particle therapy. *Journal of Instrumentation*.
- [Belosi et al., 2017] Belosi, M. F., van der Meer, R., Garcia de Acilu Laa, P., Bolsi, A., Weber, D. C., and Lomax, A. J. (2017). Treatment log files as a tool to identify treatment plan sensitivity to inaccuracies in scanned proton beam delivery. *Radiotherapy and Oncology*, 125(3):514–519.
- [Bernatowicz et al., 2018] Bernatowicz, K., Geets, X., Barragan, A., Janssens, G., Souris, K., and Sterpin, E. (2018). Feasibility of online IMPT adaptation using fast, automatic and robust dose restoration. *Physics in Medicine & Biology*, 63(8):085018.
- [Carrasco et al., 2012] Carrasco, P., Jornet, N., Latorre, A., Eudaldo, T., Ruiz, A., and Ribas, M. (2012). 3D DVH-based metric analysis versus per-beam planar analysis in IMRT pretreatment verification. *Medical Physics*.
- [Chowdhuri Z et al., 2016] Chowdhuri Z, Rohrer B, Schid F J, Faf, Gagnon-Moisan F, Koenig S, and van der Meer R (2016). The PSI water column phantom for patient-specific treatment plan verification.
- [Ford et al., 2012] Ford, E. C., Terezakis, S., Souranis, A., Harris, K., Gay, H., and Mutic, S. (2012). Quality Control Quantification (QCQ): A Tool to Measure the Value of Quality Control Checks in Radiation Oncology. *Radiation Oncology Biology*, 84(3):e263–e269.
- [Furukawa et al., 2013] Furukawa, T., Inaniwa, T., Hara, Y., Mizushima, K., Shirai, T., and Noda, K. (2013). Patient-specific QA and delivery verification of scanned ion beam at NIRS-HIMAC. *Medical Physics*, 40(12):121707.
- [Grassberger et al., 2015] Grassberger, C., Lomax, A., and Paganetti, H. (2015). Characterizing a proton beam scanning system for Monte Carlo dose calculation in patients. *Physics in Medicine and Biology*, 60(2):633–645.
- [Grevillot et al., 2011] Grevillot, L., Bertrand, D., Dessy, F., Freud, N., and Sarrut, D. (2011). A Monte Carlo pencil beam scanning model for proton treatment plan simulation using GATE/GEANT4. *Physics in Medicine and Biology*, 56(16):5203–5219.
- [Grossmann, 2007] Grossmann, M. (2007). Cancer Therapy with Protons at Paul Scherrer Institut. In *2007 15th IEEE-NPSS Real-Time Conference*, pages 1–3. IEEE.

- [Li et al., 2013] Li, H., Sahoo, N., Poenisch, F., Suzuki, K., Li, Y., Li, X., Zhang, X., Lee, A. K., Gillin, M. T., and Zhu, X. R. (2013). Use of treatment log files in spot scanning proton therapy as part of patient-specific quality assurance. *Medical Physics*, 40(Mlc):021703.
- [Li and Hsi, 2017] Li, Y. and Hsi, W. C. (2017). Analysis of measurement deviations for the patient-specific quality assurance using intensity-modulated spot-scanning particle beams. *Physics in Medicine and Biology*, 62(7):2675–2693.
- [Lomax et al., 2004a] Lomax, A. J., Böhringer, T., Bolsi, A., Coray, D., Emert, F., Goitein, G., Jermann, M., Lin, S., Pedroni, E., Rutz, H., Stadelmann, O., Timmermann, B., Verwey, J., and Weber, D. C. (2004a). Treatment planning and verification of proton therapy using spot scanning: Initial experiences. *Medical Physics*, 31(11):3150–3157.
- [Lomax et al., 2004b] Lomax, A. J., Pedroni, E., Rutz, H., and Goitein, G. (2004b). The clinical potential of intensity modulated proton therapy. *Zeitschrift für medizinische Physik*, 14(3):147–52.
- [Lomax et al., 1996] Lomax, J. A., Pedroni, E., Schaffner, B., Scheib, S., Schneider, U., Tourovsky, A., and Lomax, J. (1996). A 3D treatment planning for conformal proton therapy by spot scanning. Quantitative Imaging in Oncology.
- [Mackin et al., 2013] Mackin, D., Li, Y., Taylor, M. B., Kerr, M., Holmes, C., Sahoo, N., Poenisch, F., Li, H., Lii, J., Amos, R., Wu, R., Suzuki, K., Gillin, M. T., Zhu, X. R., and Zhang, X. (2013). Improving spot-scanning proton therapy patient specific quality assurance with HPlusQA, a second-check dose calculation engine. *Medical Physics*.
- [Meier et al., 2015] Meier, G., Besson, R., Nanz, A., Safai, S., and Lomax, A. J. (2015). Independent dose calculations for commissioning, quality assurance and dose reconstruction of PBS proton therapy. *Physics in Medicine and Biology*, 60(7):2819–2836.
- [Nelms et al., 2011] Nelms, B. E., Zhen, H., and Tomé, W. A. (2011). Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Medical physics*, 38(2):1037–44.
- [Noel et al., 2014] Noel, C. E., Santanam, L., Parikh, P. J., and Mutic, S. (2014). Process-based quality management for clinical implementation of adaptive radiotherapy. *Medical Physics*.
- [Pedroni et al., 2004] Pedroni, E., Bearpark, R., Böhringer, T., Coray, A., Duppich, J., Forss, S., George, D., Grossmann, M., Goitein, G., Hilbes, C., Jermann, M., Lin, S., Lomax, A., Negrazus, M., Schippers, M., and Kotrle, G. (2004). The PSI Gantry 2: a second generation proton scanning gantry. *ORIGINALARBEIT Z. Med. Phys*, 14:25–34.
- [Pedroni et al., 2011] Pedroni, E., Meer, D., Bula, C., Safai, S., and Zenklusen, S. (2011). Pencil beam characteristics of the next-generation proton scanning gantry of PSI: design issues and initial commissioning results. *Eur. Phys. J. Plus*, 126.
- [Raaymakers et al., 2017] Raaymakers, B. W., Jürgenliemk-Schulz, I. M., Bol, G. H., Glitzner, M., Kotte, T. J., Van Asselen, B., De Boer, J., Bluemink, J. J., Hackett, S. L., Moerland, M. A., Woodings, S. J., Wolthaus, J. W. H., Van Zijp, H. M., Philippons, M. E. P., Tijssen, R., Kok, J. G. M., De Groot-Van Breugel, E. N., Kiekebosch, I., Meijers, L. T. C., Nomden, C. N., Sikkens, G. G., Doornaert, P. A. H., Eppinga, W. S. C., Kasperts, N., Kerkmeijer, L. G. W., Tersteeg, J. H. A., Brown, K. J., Pais, B., Woodhead, P., and Lagendijk, J. J. W. (2017). First patients treated with a 1.5 T MRI- Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Physics in Medicine & Biology*.
- [Scandurra et al., 2016] Scandurra, D., Albertini, F., Van Der Meer, R., Meier, G., Weber, D. C., Bolsi, A., and Lomax, A. (2016). Assessing the quality of proton PBS treatment delivery using machine log files: Comprehensive analysis of clinical treatments delivered at PSI Gantry 2. *Physics in Medicine and Biology*.
- [Stock et al., 2017] Stock, M., Georg, D., Ableitinger, A., Zechner, A., Utz, A., Mumot, M., Kragl, G., Hopfgartner, J., Gora, J., Böhlen, T., Grevillot, L., Kuess, P., Steininger, P., Deutschmann, H., and Vatnitsky, S. (2017). The technological basis for adaptive ion beam therapy at MedAustron: Status and outlook. *Zeitschrift für medizinische Physik*.
- [Trnková et al., 2016] Trnková, P., Bolsi, A., Albertini, F., Weber, D. C., and Lomax, A. J. (2016).

Factors influencing the performance of patient specific quality assurance for pencil beam scanning IMPT fields. *Medical Physics*, 43(11):5998–6008.

- [Winterhalter et al., 2018] Winterhalter, C., Fura, E., Tian, Y., Aitkenhead, A., Bolsi, A., Dieterle, M., Fredh, A., Meier, G., Oxley, D., Siewert, D., Weber, D. C., Lomax, A., and Safai, S. (2018). Validating a Monte Carlo approach to absolute dose quality assurance for proton pencil beam scanning. *Submitted to PMB*.
- [Zhu et al., 2015] Zhu, X. R., Li, Y., Mackin, D., Li, H., Poenisch, F., Lee, A. K., Mahajan, A., Frank, S. J., Gillin, M. T., Sahoo, N., and Zhang, X. (2015). Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. *Cancers*, 7(2):631–647.
- [Zhu et al., 2011] Zhu, X. R., Poenisch, F., Song, X., Johnson, J. L., Ciangaru, G., Taylor, M. B., Lii, M., Martin, C., Arjomandy, B., Lee, A. K., Choi, S., Nguyen, Q. N., Gillin, M. T., and Sahoo, N. (2011). Patient-specific quality assurance for prostate cancer patients receiving spot scanning proton therapy using single-field uniform dose. *International Journal of Radiation Oncology Biology Physics*.